Performance Measurement and Visualization on the Cray XT

Luiz DeRose
Programming Environment Director
Cray Inc.
ldr@cray.com

ORNL
October 15, 2008

The Cray Tools Strategy

- Must be easy and flexible to use
  - Automatic program instrumentation
    - no source code or makefile modification needed
- Integrated performance tools solution
  - Multiple platforms
  - Multiple functionality
    - Measurements of user functions, MPI, I/O, memory, & math SW
    - HW Counters support
- Close interaction with user for feedback targeting functionality enhancements

Cray Performance Analysis Infrastructure

- CrayPat
  - pat_build: Utility for application instrumentation
    - No source code modification required
  - run-time library for measurements
    - transparent to the user
  - pat_report:
    - Performance reports
    - Performance visualization file
  - pat_help
    - Interactive performance tool help utility

- Cray Apprentice^2
  - Graphical performance analysis and visualization tool
  - Can be used off-line on Linux system

Performance Data Collection

- Two dimensions
  - When performance collection is triggered
    - Externally (asynchronous)
      - Sampling
        - » Timer interrupt
        - » Hardware counters overflow
    - Internally (synchronous)
      - Code instrumentation
        - » Event based
        - » Automatic or manual instrumentation
  - How performance data is recorded
    - Profile ::= Summation of events over time
      - run time summarization (functions, call sites, loops, …)
    - Trace file ::= Sequence of events over time

Performance Analysis with Cray Tools

- Important performance statistics:
  - Top time consuming routines
  - Load balance across computing resources
  - Communication overhead
  - Cache utilization
  - FLOPS
  - Vectorization (SSE instructions)
  - Ratio of computation versus communication

Application Instrumentation with pat_build

- No source code or makefile modification required
  - Automatic instrumentation at group (function) level
    - Groups: mpi, io, heap, math SW, …
- Performs link-time instrumentation
  - Requires object files
  - Instruments optimized code
  - Generates stand-alone instrumented program
  - Preserves original binary
  - Supports sample-based and event-based instrumentation
- pat_build [-d dirfile] [-D directive] [-f] [-g tracegroup] [-n] [-O ofile] [-o instr_program] [-t tracefile] [-T tracefunc] [-u] [-V] [-v] [-w] [-z] program [instr_program]

-g tracegroup

- biolibs — Cray Bioinformatics library routines
- blas — Basic Linear Algebra subprograms
- heap — dynamic heap
- io — includes stdio and sysio groups
- lapack — Linear Algebra Package
- math — ANSI math
- mpi — MPI
- omp — OpenMP API
- omp-rtl — OpenMP runtime library (not supported on Catamount)
- pthreads — POSIX threads (not supported on Catamount)
- shmem — SHMEM
- stdio — all library functions that accept or return the FILE* construct
- sysio — I/O system calls
- system — system calls

Running the Instrumented Application

- MUST run on Lustre (/tmp/work/…)

- Runtime environment variables
  - Optional timeline view of program available
    - export PAT_RT_SUMMARY=0
    - View trace file with Cray Apprentice^2
  - Number of files used to store raw data:
    - 1 file created for program with 1 – 256 processes
    - $\sqrt{n}$ files created for program with 257 – *n* processes
    - Ability to customize with PAT_RT_EXPFILE_MAX
  - Request hardware performance counter information:
    - export PAT_RT_HWPC=<HWPC Group>
    - Can specify events or predefined groups

pat_report

- Performs data conversion
  - Combines information from binary with raw performance data
- Generates text report of performance results
- Formats data for input into Cray Apprentice^2

  - pat_report [-V] [-i dir|instrprog] [-o output_file] [-O keyword] [-C 'table_caption'] [-d d-opts] [-b b-opts] [-s key=value] [-H] [-P] [-T] [-z] data_directory | data_file.xf

What is the ".ap2" File?

- The ".ap2" file is a self contained compressed performance file
  - Normally it is about 5 times smaller than the ".xf" file
  - Contains the information needed from the application binary
    - Can be reused, even if the application binary is no longer available or if it was rebuilt
  - It is the only input format accepted by Cray Apprentice^2
- The "ap2" file is generated by default when executing pat_report

Automatic Profiling Analysis (APA)

1. Load CrayPat & Cray Apprentice^2 module files
   - % module load xt-craypat apprentice2
2. Build application
   - % make clean
   - % make
3. Instrument application for automatic profiling analysis
   - % pat_build **–O apa** a.out
     - You should get an instrumented program a.out+pat
4. Run application to get top time consuming routines
   - Remember to modify <script> to run *a.out+pat*
   - Remember to run on Lustre
   - % aprun … *a.out+pat* *(or* qsub <pat script>)
     - You should get a performance file (“<sdatafile>.xf”) or multiple files in a directory <sdatadir>
5. Generate .apa file
   - % pat_report –o *my_sampling_report* [<sdatafile>.xf | <sdatadir>]
     - creates a report file & an automatic profile analysis file <apafile>.apa

APA File Example

```
#  You can edit this file, if desired, and use it
#  to reinstrument the program for tracing like this:
#
#          pat_build -O mhd3d.Oapa.x+4125-401sdt.apa
#
#  These suggested trace options are based on data from:
#
#       /home/crayadm/ldr/mhd3d/run/mhd3d.Oapa.x+4125-401sdt.ap2, /home/
           crayadm/ldr/mhd3d/run/mhd3d.Oapa.x+4125-401sdt.xf

# ----------------------------------------------------------------------

#       HWPC group to collect by default.

  -Drtenv=PAT_RT_HWPC=0  # Summary with instructions metrics.

# ----------------------------------------------------------------------

#       Libraries to trace.

  -g mpi

# ----------------------------------------------------------------------

#       User-defined functions to trace, sorted by % of samples.
#       Limited to top 200. A function is commented out if it has < 1%
#       of samples, or if a cumulative threshold of 90% has been reached,
#       or if it has size < 200 bytes.

  -w  # Enable tracing of user-defined functions.
      # Note: -u should NOT be specified as an additional option.
```

```
# 43.37%  99659 bytes
        -T mlwxyz_

# 16.09%  17615 bytes
        -T half_

#  7.98%  12666 bytes
        -T full_

#  6.82%  6846 bytes
        -T artv_

. . .

#  1.29%  5352 bytes
        -T currenh_

#  1.03%  25294 bytes
        -T bndbo_

#  Functions below this point account for less than 10% of samples.


#  1.03%  31240 bytes
#         -T bndto_

#  0.51%  11169 bytes
#         -T bncto_

. . .

# ----------------------------------------------------------------------

  -o mhd3d.x+apa                # New instrumented program.

  /work/crayadm/ldr/mhd3d/mhd3d.x  # Original program.
```

Steps to collecting performance data

6. Look at <apafile>.apa file
   - Verify if additional instrumentation is wanted

7. Instrument application for further analysis (a.out+apa)
   - % pat_build **–O <apafile>.apa**
     - You should get an instrumented program a.out+apa

8. Run application
   - Remember to modify <script> to run *a.out+apa*
   - % aprun … *a.out+apa* *(or* qsub <apa script>)
     - You should get a performance file (“<datafile>.xf”) or multiple files in a directory <datadir>

9. Create text report
   - % pat_report –o *my_text_report.txt* [<datafile>.xf | <datadir>]
     - Will generate a compressed performance file (<datafile>.ap2)

10. View results in text (my_text_report.txt) and/or with Cray Apprentice^2
    - % app2 <*datafile*>.ap2

CrayPat API - for fine grain instrumentation

- Fortran

```
include "pat_apif.h"
…
call PAT_region_begin(id, "label", ierr)
do i = 1,n
…
enddo
call PAT_region_end(id, ierr)
```

- C

```
include <pat_api.h>
…
ierr = PAT_region_begin(id, "label");
< code segment >
ierr = PAT_region_end(id);
```

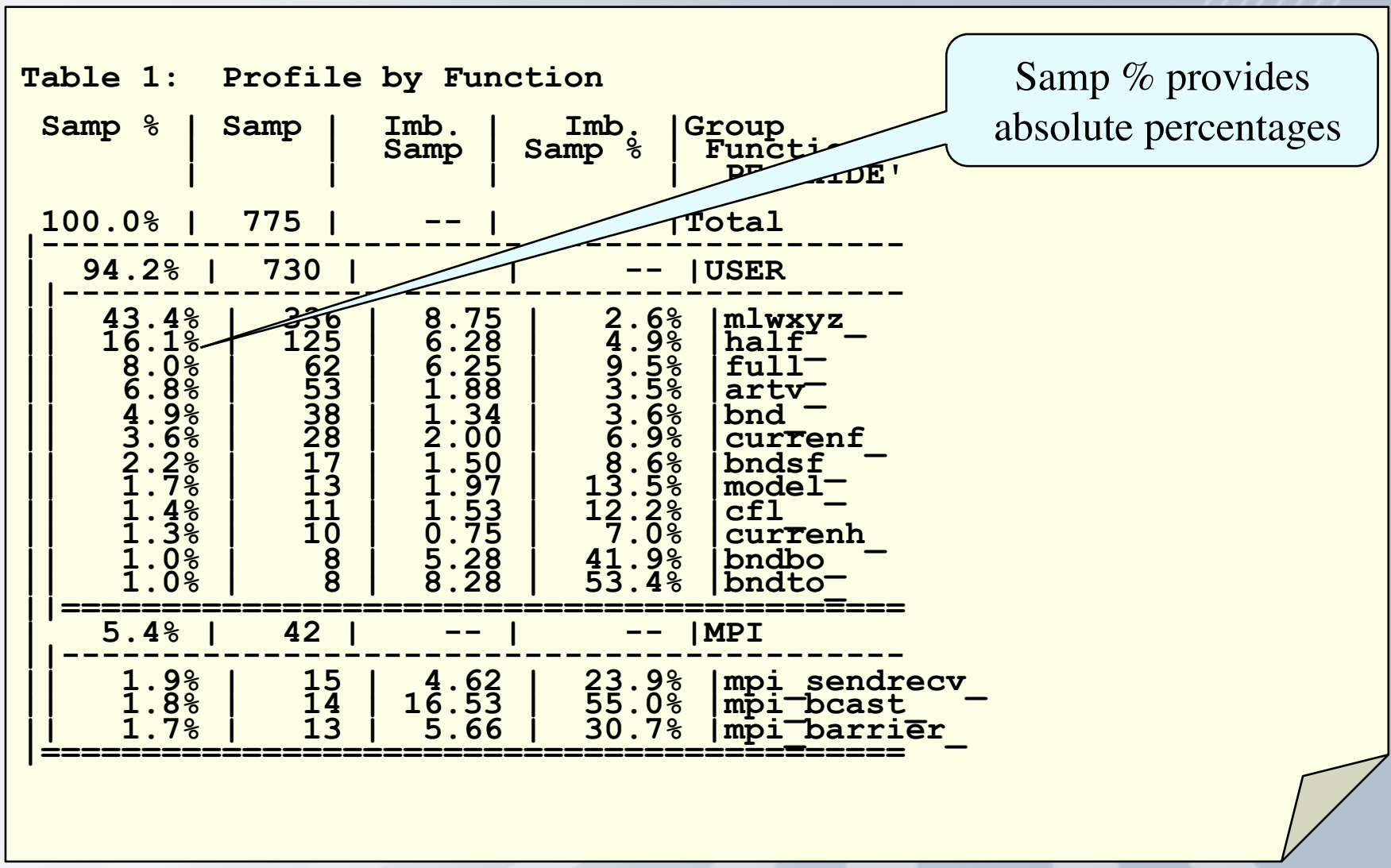
Pat_report Output

```
CrayPat/X:  Version 4.3.0 Revision 1803 (xf 1649)  06/10/08 12:01:46
Number of PEs (MPI ranks):      32
Number of Threads per PE:       1
Number of Cores per Processor:  4
Execution start time:   Mon Jun 23 14:07:31 2008
System type and speed:   x86_64  2100 MHz
Current path to data file:
  /home/crayadm/ldr/mhd3d/run/mhd3d.Oapa.x+4681-295sdt.ap2  (RTS)
  /home/crayadm/ldr/mhd3d/run/mhd3d.Oapa.x+4681-295sdt.xf   (RTS)
Notes:
    Sampling interval was 10000 microseconds (100.0/sec)
  BSD timer type was ITIMER_PROF


  Trace option suggestions have been generated into a separate file
  from the data in the next table.  You can examine the file, edit it
  if desired, and use it to reinstrument the program like this:

            pat_build -O mhd3d.Oapa.x+4681-295sdt.apa
  (To see the list, specify:  -s traced_functions=show)
...
```

Sampling Output (Notes for Table 1)

```
Notes for table 1:

  Table option:
    -O samp_profile
  Options implied by table option:
    -d sa%@0.95,sa,imb_sa,imb_sa% -b gr,fu,pe=HIDE

  Options for related tables not shown by default:
    -O samp_profile+src

  The Total value for Samp is the sum of the Group values.
  The Group value for Samp is the sum of the Function values.
  The Function value for Samp is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with Samp% > 0.95.
    (To set thresholds to zero, specify:  -T)

  Percentages at each level are of the Total for the program.
    (For percentages relative to next level up, specify:
      -s percent=r[elative])
```

Sampling Output (Table 1)



Sampling Output (Notes for Table 2)

```
Notes for table 2:

  Table option:
    -O samp_profile+src
  Options implied by table option:
    -d sa%@0.95,sa,imb_sa,imb_sa% -b gr,fu,so,li,pe=HIDE

  Options for related tables not shown by default:
    -O samp_profile

  The Total value for Samp is the sum of the Group values.
  The Group value for Samp is the sum of the Function values.
  The Function value for Samp is the sum of the Source values.
  The Source value for Samp is the sum of the Line values.
  The Line value for Samp is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with Samp% > 0.95.
    (To set thresholds to zero, specify:  -T)

  Percentages at each level are of the Total for the program.
    (For percentages relative to next level up, specify:
      -s percent=r[elative])
```

Sampling Output (Table 2)

```
Table 2:  Profile by Group, Function, and Line

 Samp % | Samp |  Imb.  |   Imb.  |Group
        |      |  Samp  | Samp %  | Function
        |      |        |         |  Source
        |      |        |         |   Line
        |      |        |         |    PE='HIDE'

 100.0% |  777 |    -- |     -- |Total
|-----------------------------------------
|  94.2% |  732 |    -- |     -- |USER
|-----------------------------------------
|   43.4% |  337 |    -- |      -- |mlwxyz
3|         |      |       |          | ldr/mhd3d/src/mlwxyz.f
||||-------------------------------------
4|||    2.1% |   16 |  1.47 |    8.9% |line.39
4|||    2.8% |   22 |  2.25 |    9.7% |line.78
4|||    1.2% |    9 |  1.09 |   11.3% |line.116
4|||    1.4% |   11 |  1.22 |   10.5% |line.129
4|||    2.2% |   17 |  2.12 |   11.5% |line.139
. . .
4|||    2.7% |   21 |  0.84 |    4.0% |line.568
4|||    1.3% |   10 |  1.72 |   14.8% |line.604
4|||    2.4% |   19 |  0.72 |    3.7% |line.634
||||=========================================
||||=========================================
|   16.1% |  125 |    -- |      -- |half
3|         |      |       |          | ldr/mhd3d/src/half.f
||||-------------------------------------
4|||    5.4% |   42 |  6.41 |   13.8% |line.28
4|||   10.7% |   83 |  5.91 |    6.9% |line.40
||||=========================================
|    8.0% |   62 |    -- |      -- |full
3|         |      |       |          | ldr/mhd3d/src/full.f
||||-------------------------------------
4|||    8.0% |   62 |  6.31 |    9.6% |line.22
||||=========================================
. . .
||=========================================
|    5.4% |   42 |    -- |     -- |MPI
||----------------------------------------
||    1.9% |   15 |  4.62 |  23.9% |mpi_sendrecv_
||||-------------------------------------
||||=========================================
||    1.8% |   14 | 16.53 |  55.0% |mpi_bcast
||    1.7% |   13 |  5.66 |  30.7% |mpi_barrier_
|==========================================
```

Sampling Output (Table 3)

```
Notes for table 3:

  Table option:
    -O program_time
  Options implied by table option:
    -d pt -b pe=[mmm]

  The Total value for Process Time is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)


Table 3:  Program Wall Clock Time

   Process |PE[mmm]
      Time |

 13.926882 |Total
|-----------------
| 14.093036 |pe.0
| 13.924961 |pe.13
| 13.744948 |pe.14
|=================
```

Table 1: Flat Profile (Default)

```
Notes for table 1:

  Table option:
    -O profile
  Options implied by table option:
    -d ti%@0.95,ti,imb_ti,imb_ti%,tr -b gr,fu,pe=HIDE
  Other options:
    -H

  Options for related tables not shown by default:
    -O profile_pe.th           -O callers
    -O profile_th_pe           -O callers+src
    -O profile+src             -O calltree
    -O load_balance            -O calltree+src

  The Total value for each of Time, Calls is the sum of the Group values.
  The Group value for each of Time, Calls is the sum of the Function values.
  The Function value for each of Time, Calls is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with Time% > 0.95.
    (To set thresholds to zero, specify:  -T)

  Percentages at each level are of the Total for the program.
    (For percentages relative to next level up, specify:
      -s percent=r[elative])
```

Table 1: Flat Profile (Continuation)

```
Table 1:  Profile by Function Group and Function

 Time % |        Time |Imb. Time |   Imb. | Calls |Group
        |             |          | Time % |       | Function
        |             |          |        |       |  PE='HIDE'

 100.0% | 104.593634 |       -- |     -- | 22649 |Total
|-----------------------------------------------------------
|  71.0% |  74.230520 |       -- |     -- | 10473 |MPI
||----------------------------------------------------------
||  69.7% |  72.905208 | 0.508369 |   0.7% |   125 |mpi_allreduce_
||   1.0% |   1.050931 | 0.030042 |   2.8% |    94 |mpi_alltoall_
||==========================================================
|  25.3% |  26.514029 |       -- |     -- |    73 |USER
||----------------------------------------------------------
||  16.7% |  17.461110 | 0.329532 |   1.9% |    23 |selfgravity_
||   7.7% |   8.078474 | 0.114913 |   1.4% |    48 |ffte4_
||==========================================================
|   2.5% |   2.659429 |       -- |     -- |   435 |MPI_SYNC
||----------------------------------------------------------
||   2.1% |   2.207467 | 0.768347 |  26.2% |   172 |mpi_barrier_(sync)
||==========================================================
|   1.1% |   1.188998 |       -- |     -- | 11608 |HEAP
||----------------------------------------------------------
||   1.1% |   1.166707 | 0.142473 |  11.1% |  5235 |free
|==========================================================
```

Table 2: Load Balance

Table 2: Load Balance with MPI Sent Message Stats

Time %	Time	Sent Msg Count	Sent Msg Total Bytes	Avg Sent Msg Size	Group PE[mmm]
100.0%	104.628869	2041	9272032	4542.89	Total
71.0%	74.246813	2041	9272032	4542.89	MPI
1.1%	74.820272	2332	10745856	4608.00	pe.41
1.1%	74.149777	2332	10745856	4608.00	pe.25
1.1%	73.934590	2332	10745856	4608.00	pe.19
25.3%	26.514143	--	--	--	USER
0.4%	26.960079	--	--	--	pe.22
0.4%	26.367891	--	--	--	pe.37
0.4%	26.212226	--	--	--	pe.48
2.5%	2.660105	--	--	--	MPI_SYNC
0.1%	3.492460	--	--	--	pe.34
0.0%	2.625068	--	--	--	pe.36
0.0%	1.683962	--	--	--	pe.15
1.2%	1.207056	--	--	--	HEAP
0.0%	1.346021	--	--	--	pe.6
0.0%	1.215907	--	--	--	pe.55
0.0%	0.960383	--	--	--	pe.49

Table 3: MPI Send Stats by Caller

```
Notes for table 3:

  Table option:
    -O mpi_callers
  Options implied by table option:
    -d sm,sc@,mb1..7 -b fu,ca,pe=[mmm]

  Options for related tables not shown by default:
    -O mpi_dest_bytes          -O mpi_dest_counts

  The Total value for each data item is the sum of the Function values.
  The Function value for each data item is the sum of the Caller values.
  The Caller value for each data item is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with Sent Msg Count > 0.


Table 3:  MPI Sent Message Stats by Caller

 Sent Msg |  Sent | 4KB<= |Function
    Total |   Msg | MsgSz | Caller
    Bytes | Count | <64KB |  PE[mmm]
          |       | Count |

  9272032 |  2041 |  2041 |Total
|-----------------------------------
|  9272032 |  2041 |  2041 |mpi_isend_
|          |       |       | mpi_shift_
|||----------------------------------
3|| 10745856 |  2332 |  2332 |pe.33
3|| 10745856 |  2332 |  2332 |pe.38
3||  3829760 |   880 |   880 |pe.63
|===================================
```

Table 5: Heap Statistics

```
Notes for table 5:

  Table option:
    -O heap_hiwater
  Options implied by table option:
    -d am@,ub,ta,ua,tf,nf,ac,ab -b pe=[mmm]
  Other options:
    -H

  The Total value for each data item is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with Tracked Heap HiWater MBytes > 0.


Table 5:  Heap Stats during Main Program

 Tracked |  Total | Total | Tracked | Tracked |PE[mmm]
    Heap | Allocs | Frees | Objects |  MBytes |
 HiWater |        |       |     Not |     Not |
  MBytes |        |       |  Freed  |   Freed |

 139.669 |   6372 |  5235 |    1137 |   9.671 |Total
|-------------------------------------------------------
| 140.946 |   6389 |  4989 |    1400 |  10.999 |pe.38
| 139.645 |   5983 |  4892 |    1091 |   9.641 |pe.14
| 138.758 |   5695 |  4656 |    1039 |   8.754 |pe.63
|=======================================================
```

Table 6: Heap Leaks

```
Notes for table 6:

  Table option:
    -O heap_leaks
  Options implied by table option:
    -d lb%@1,lb@0.0005,lc -b ca,pe=[mmm]

  The Total value for each of Tracked Objects Not Freed, Tracked MBytes Not
    Freed is the sum of the Caller values.
  The Caller value for each of Tracked Objects Not Freed, Tracked MBytes Not
    Freed is the avg of the PE values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with:
    Tracked MBytes Not Freed% > 1
    Tracked MBytes Not Freed > 0.0005    (To set thresholds to zero, specify:   -T)

Table 6:  Heap Leaks during Main Program
 Tracked | Tracked | Tracked |Caller
  MBytes |  MBytes | Objects | PE[mmm]
     Not |     Not |     Not |
 Freed % |   Freed |   Freed |

  100.0% |   0.010 |       1 |Total
|-----------------------------------
|   95.7% |   0.010 |       1 |main
||----------------------------------
||    4.2% |   0.010 |       1 |pe.11
||    4.2% |   0.010 |       1 |pe.6
||    4.2% |   0.010 |       1 |pe.5
|===================================
```

Table 7: I/O (Read) Statistics

```
Notes for table 7:

  Table option:
    -O read_stats
  Options implied by table option:
    -d rt,rb,rR,rd@,rC -b fi,pe=[mmm],fd

  The Total value for each data item is the sum of the File Name values.
  The File Name value for each of Read B/Call, Read Time, Reads is the avg of
    the PE values.
  The File Name value for each of Read Rate MB/sec, Read MB is the sum of the
    PE values.
  The PE value for each data item is the sum of the File Desc values.
    (To specify different aggregations, see:  pat_help report options s1)

  This table shows only lines with Reads > 0.


Table 7:  File Input Stats by Filename

  Read |  Read MB |Read Rate | Reads |   Read |File Name
  Time |          |   MB/sec |       | B/Call | PE[mmm]
       |          |          |       |        |  File Desc

 0.000 | 0.000065 | 0.925430 |     3 |  22.67 |Total
|-----------------------------------------------------
| 0.000 | 0.000065 | 0.925430 |     3 |  22.67 |input
||----------------------------------------------------
|| 0.002 | 0.000065 | 0.038560 |    68 |   1.00 |pe.0
3|       |          |          |       |        | fd.12
|| 0.000 |       -- |       -- |    -- |     -- |pe.6
|| 0.000 |       -- |       -- |    -- |     -- |pe.5
|====================================================
```

Table 8: I/O (Write) Statistics

```
Table 8:  File Output Stats by Filename

 Write | Write MB | Write Rate | Writes |  Write |File Name
  Time |          |     MB/sec |        | B/Call | PE[mmm]
       |          |            |        |        |  File Desc

 0.000 | 0.002298 |  14.088269 |     12 | 200.83 |Total
|-------------------------------------------------------------
| 0.000 | 0.000298 |   2.070438 |      1 | 312.00 |stderr
||------------------------------------------------------------
|| 0.000 | 0.000012 |   0.026614 |      1 |  13.00 |pe.17
3|       |          |            |        |        | fd.2
|| 0.000 | 0.000012 |   0.253220 |      1 |  13.00 |pe.7
3|       |          |            |        |        | fd.2
|| 0.000 | 0.000012 |   2.840267 |      1 |  13.00 |pe.0
3|       |          |            |        |        | fd.2
||============================================================
| 0.000 | 0.002001 | 102.986588 |     11 | 190.73 |stdout
||------------------------------------------------------------
|| 0.000 | 0.002001 |   4.291078 |    269 |   7.80 |pe.0
3|       |          |            |        |        | fd.1
|| 0.000 |       -- |         -- |     -- |     -- |pe.6
|| 0.000 |       -- |         -- |     -- |     -- |pe.5
|=============================================================
```

Pat_report –O options

```
% pat_report -O help
pat_report: Help for -O option:
Available option values are in left column, a prefix can be specified:
  ct                        -O calltree
  defaults                  Tables that would appear by default.
  heap                      -O heap_program,heap_hiwater,heap_leaks
  io                        -O read_stats,write_stats
  lb                        -O load_balance
  load_balance              -O lb_program,lb_group,lb_function
  mpi                       -O mpi_callers
  ---
  callers                   Profile by Function and Callers
  callers+src               Profile by Function and Callers, with Line Numbers
  calltree                  Function Calltree View
  calltree+src              Calltree View with Callsite Line Numbers
  heap_hiwater              Heap Stats during Main Program
  heap_leaks                Heap Leaks during Main Program
  heap_program              Heap Usage at Start and End of Main Program
  hwpc                      HW Performance Counter Data
  load_balance_function     Load Balance across PE's by Function
  load_balance_group        Load Balance across PE's by FunctionGroup
  load_balance_program      Load Balance across PE's
  load_balance_sm           Load Balance with MPI Sent Message Stats
  loops                     Loop Stats from -hprofile_generate
  mpi_callers               MPI Sent Message Stats by Caller
  mpi_dest_bytes            MPI Sent Message Stats by Destination PE
  mpi_dest_counts           MPI Sent Message Stats by Destination PE
  mpi_rank_order            Suggested MPI Rank Order
  mpi_sm_rank_order         Sent Message Stats and Suggested MPI Rank Order
  pgo_details               Loop Stats detail from -hprofile_generate
  profile                   Profile by Function Group and Function
  profile+src               Profile by Group, Function, and Line
  profile_pe.th             Profile by Function Group and Function
  profile_pe_th             Profile by Function Group and Function
  profile_th_pe             Profile by Function Group and Function
  program_time              Program Wall Clock Time
  read_stats                File Input Stats by Filename
  samp_profile              Profile by Function
  samp_profile+src          Profile by Group, Function, and Line
  thread_times              Program Wall Clock Time
  write_stats               File Output Stats by Filename
```

Callers Profile (Bottom Up)

```
 Notes for table 1:

  Table option:
    -O callers
  Options implied by table option:
    -d ti%@0.95,ti,tr -b gr,fu,ca,pe=HIDE
  Other options:
    -H

Table 1:  Profile by Function and Callers

 Time % |        Time | Calls |Group
        |             |       | Function
        |             |       |  Caller
        |             |       |   PE='HIDE'

 100.0% | 104.593634 | 22651 |Total
|-----------------------------------------
|  71.0% |  74.230520 | 10473 |MPI
||----------------------------------------
||  69.7% |  72.905208 |   125 |mpi_allreduce_
|||---------------------------------------
3||  47.0% |  49.206835 |    48 |ffte4_
||||--------------------------------------
4|||  46.1% |  48.193285 |    47 |ffte4_(exclusive)
4|||   1.0% |   1.013550 |     1 |selfgravity_
||||======================================
3||  22.7% |  23.694255 |    23 |selfgravity_
|||======================================
||   1.0% |   1.050931 |    94 |mpi_alltoall_
3|        |            |       | pztrans_
||=========================================
```

Callers Profile – MPI (Cont.)

```
Table 1:  Profile by Function and Callers

 Time % |        Time | Calls |Group
        |             |       | Function
        |             |       |  Caller
        |             |       |   PE='HIDE'

 100.0% | 103.200155 | 10981 |Total
|-----------------------------------------
|  72.2% |  74.494482 | 10473 |MPI
||----------------------------------------
||  71.1% |  73.332263 |   125 |mpi_allreduce_
|||---------------------------------------
3||  47.9% |  49.425113 |    48 |ffte4_
||||-------------------------------------
4|||  46.9% |  48.386489 |    47 |ffte4_(exclusive)
4|||   1.0% |   1.038624 |     1 |selfgravity_
||||=====================================
3||  23.2% |  23.904200 |    23 |selfgravity_
||=======================================
|  25.5% |  26.347519 |    73 |USER
||---------------------------------------
||  16.8% |  17.304612 |    23 |selfgravity_
|||-------------------------------------
3||   7.9% |   8.201878 |    11 |firststep_
3||   7.9% |   8.201270 |    11 |secondstep_
|||=======================================
||   7.8% |   8.094927 |    48 |ffte4_
3|        |            |       | selfgravity_
||=======================================
|   2.3% |   2.358154 |   435 |MPI_SYNC
||--------------------------------------
||   2.0% |   2.060709 |   172 |mpi_barrier_(sync)
|||-------------------------------------
3||   2.0% |   2.025665 |    96 |ffte4_
||||-------------------------------------
4|||   1.1% |   1.172514 |    49 |selfgravity_
|=========================================

|================================================
```

Call Tree Profile (Top Down)

```
Table 1:  Function Calltree View

 Time % |      Time | Calls |Calltree
        |           |       | PE='HIDE'

 100.0% | 12.306348 |  1396 |Total
|-----------------------------------
| 100.0% | 12.302515 |   664 |main
||----------------------------------
||  99.4% | 12.226615 |   661 |MAIN_
|||---------------------------------
3||  80.5% |  9.911106 |   390 |mlwxyz_
||||--------------------------------
4|||  43.8% |  5.394583 |    10 |mlwxyz_(exclusive)
4|||  14.9% |  1.839600 |    80 |half_
4|||   8.0% |  0.985683 |    80 |full_
4|||   7.0% |  0.859738 |    80 |artv_
4|||   4.6% |  0.560140 |    70 |currenf_
||||-------------------------------
5||||   4.2% |  0.514998 |    10 |currenf_(exclusive)
||||===============================
4|||   2.2% |  0.271362 |    70 |currenh_
||||-------------------------------
5||||   1.9% |  0.231544 |    10 |currenh_(exclusive)
||||===============================
3||  11.4% |  1.406021 |    99 |bnd_
||||--------------------------------
4|||   7.1% |  0.870392 |    11 |bnd_(exclusive)
4|||   2.1% |  0.259125 |    11 |bndsf_
4|||   1.6% |  0.201370 |    22 |bndbo_
||||-------------------------------
5||||   1.0% |  0.127880 |    11 |bndbo_(exclusive)
||||===============================
3||   3.2% |  0.393540 |     3 |mpi_bcast_(sync)
3||   2.0% |  0.245981 |     1 |model_
||||-------------------------------
4|||   2.0% |  0.245981 |     1 |model_(exclusive)
||||===============================
3||   1.5% |  0.188091 |    33 |cfl_
||||-------------------------------
4|||   1.5% |  0.184726 |    11 |cfl_(exclusive)
```

Callers Profile with Line Numbers

```
Notes for table 1:

  Table option:
    -O calltree+src
  Options implied by table option:
    -d ti%@0.95,ti,tr -b ct,pe=HIDE -s show_ca='fu,so,li' \
    -s source_limit='1'
  Other options:
    -H

Table 1:  Calltree View with Callsite Line Numbers

 Time % |      Time | Calls |Calltree
        |           |       | PE='HIDE'

 100.0% | 12.381493 |  1155 |Total
|-----------------------------------
| 100.0% | 12.380694 |   663 |main:...:line.0
||----------------------------------
||  91.8% | 11.370849 |   560 |MAIN_:mhdmain.f:line.368
3||  43.5% |  5.387737 |    10 |mlwxyz_:mlwxyz.f:line.2
3||   6.9% |  0.849967 |    80 |mlwxyz_:mlwxyz.f:line.604
|||-----------------------------------
4|||       |           |       |artv_:artv.f:line.2
|||===================================
3||   6.5% |  0.806207 |    80 |mlwxyz_:mlwxyz.f:line.283
|||-----------------------------------
4||   4.2% |  0.514861 |    10 |currenf_:currenf.f:line.2
4||   2.0% |  0.247048 |    10 |half_:half.f:line.2
|||===================================
3||   6.5% |  0.798654 |    10 |bnd_:bnd.f:line.2
3||   3.9% |  0.487620 |    80 |bnd_:bnd.f:line.15
|||-----------------------------------
4||   1.9% |  0.234759 |    10 |bndsf_:bndsf.f:line.2
|||===================================
3||   2.2% |  0.271831 |    70 |mlwxyz_:mlwxyz.f:line.39
|||-----------------------------------
4||   1.9% |  0.232804 |    10 |currenh_:currenh.f:line.2
|||===================================
3||   2.1% |  0.253929 |    10 |mlwxyz_:mlwxyz.f:line.253
4||   2.0% |  0.252611 |    10 | half_:half.f:line.2
3||   2.0% |  0.249683 |    10 |mlwxyz_:mlwxyz.f:line.129
4||   2.0% |  0.249683 |    10 | half_:half.f:line.2
3||   2.0% |  0.249683 |    10 |mlwxyz_:mlwxyz.f:line.273
```

Load Balancing Function per PE

```
Notes for table 1:

  High level option:  -O load_balance_program
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,pe

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 1:  Load Balance across PE's

 Time % |   Cum. |     Time |  Calls |Experiment=1
        | Time % |          |        |PE

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-------------------------------------------------
|   2.1% |   2.1% | 3.823080 |   7160 |pe.0
|   2.1% |   4.2% | 3.799148 |  13753 |pe.8
| ...
|   2.1% |  97.9% | 3.796151 |   7683 |pe.5
|   2.1% | 100.0% | 3.796144 |  10431 |pe.29
|=================================================
```

Table 3: LB Across PE's by Function

```
Notes for table 3:

  High level option:  -O load_balance_function
  Low level options:  -d ti%@0.05,cum_ti%,ti,tr -b exp,gr,fu,pe

  This table shows only lines with Time% > 0.05.

  Percentages at each level are relative
    (for absolute percentages, specify:  -s percent=a).


Table 3:  Load Balance across PE's by Function

 Time % |   Cum. |     Time |  Calls |Experiment=1
        | Time % |          |        |Group
        |        |          |        | Function
        |        |          |        |  PE

 100.0% | 100.0% | 3.798177 | 579653 |Total
|-----------------------------------------------
|  70.9% |  70.9% | 2.692783 | 245380 |USER
||----------------------------------------------
||  97.1% |  97.1% | 2.615916 |    576 |sweep_
|||---------------------------------------------
|||   2.2% |   2.2% | 2.753279 |     12 |pe.0
|||   2.1% |   4.3% | 2.654725 |     12 |pe.5
||| . . .
|||   2.0% |  98.0% | 2.525587 |     12 |pe.43
|||   2.0% | 100.0% | 2.523325 |     12 |pe.37
|||=============================================
 . . .
||==============================================
||   0.4% |  99.2% | 0.010300 | 118080 |snd_real_
|||---------------------------------------------
|||   2.4% |   2.4% | 0.011699 |   2880 |pe.26
|||   2.3% |   4.7% | 0.011475 |   2880 |pe.27
||| . . .
|||   1.5% |  98.6% | 0.007266 |   1440 |pe.0
|||   1.4% | 100.0% | 0.006907 |   1440 |pe.5
|||=============================================
```

Table 3 (Cont.)

```
||=================================================
|  28.8% |  99.7% | 1.092307 | 238224 |MPI
||-------------------------------------------------
||  76.1% |  76.1% | 0.831311 | 118080 |mpi_recv_
|||-----------------------------------------------
|||   2.7% |   2.7% | 1.066077 |   1440 |pe.47
|||   2.6% |   5.3% | 1.034307 |   2160 |pe.41
||| . . .
|||   1.8% |  98.6% | 0.700970 |   2160 |pe.1
|||   1.4% | 100.0% | 0.573420 |   1440 |pe.0
|||===============================================
. . .
||=================================================
|   0.2% |  99.8% | 0.007329 |  95597 |HEAP
||-------------------------------------------------
||  61.1% |  61.1% | 0.004481 |  47861 |malloc
|||-----------------------------------------------
|||   2.7% |   2.7% | 0.005884 |   1242 |pe.12
|||   2.6% |   5.4% | 0.005658 |   1226 |pe.19
||| . . .
|||   1.3% |  99.5% | 0.002827 |    618 |pe.34
|||   0.5% | 100.0% | 0.001164 |    417 |pe.0
|||===============================================
||  38.9% | 100.0% | 0.002848 |  47735 |free
|||-----------------------------------------------
|||   2.7% |   2.7% | 0.003748 |   1422 |pe.37
|||   2.7% |   5.5% | 0.003706 |   1469 |pe.43
||| . . .
|||   1.4% |  99.3% | 0.001867 |    616 |pe.34
|||   0.7% | 100.0% | 0.000896 |    385 |pe.0
||=================================================
|   0.2% | 100.0% | 0.005758 |    452 |IO
||-------------------------------------------------
||  81.3% |  81.3% | 0.004679 |    309 |fwrite
|||-----------------------------------------------
|||  34.3% |  34.3% | 0.077141 |    262 |pe.0
|||   2.1% |  36.4% | 0.004615 |      1 |pe.8
||| . . .
```

Documentation

- The pat_help utility is an interactive viewer used to access information about and examples of using CrayPat
  - pat_help [topic [subtopic...]]

- See also man pages:
  - intro_craypat
  - pat_build
  - pat_report
  - pat_help
  - pat_hwpc
  - hwpc
  - papi_counters
  - app2

pat_help Example

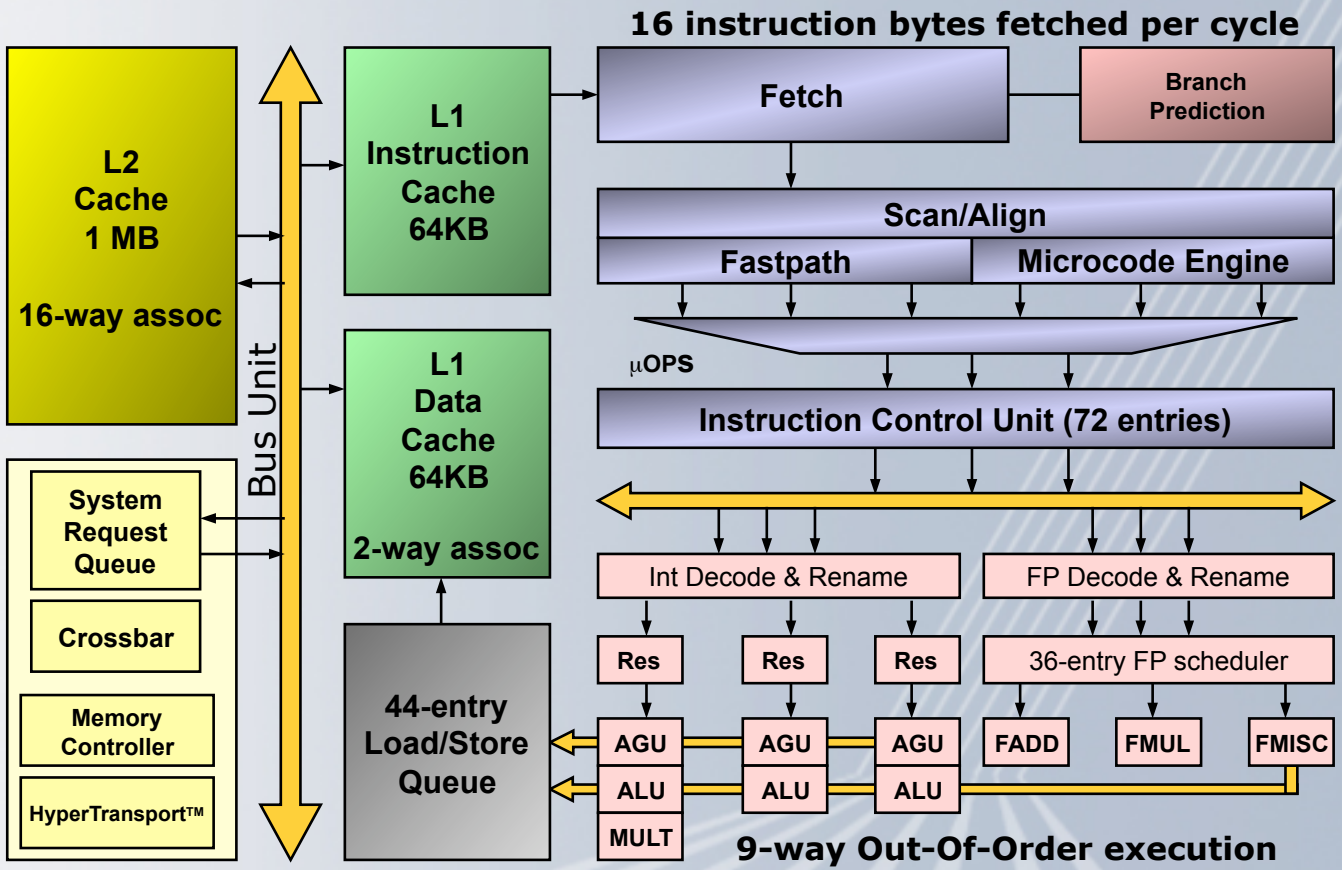
```
% pat_help

    The top level CrayPat/X help topics are listed below.
    A good place to start is:

        overview

    If a topic has subtopics, they are displayed under the heading
    "Additional topics", as below.  To view a subtopic, you need
    only enter as many initial letters as required to distinguish
    it from other items in the list.  To see a table of contents
    including subtopics of those subtopics, etc., enter:

        toc

    To produce the full text corresponding to the table of contents,
    specify "all", but preferably in a non-interactive invocation:

        pat_help all . > all_pat_help
        pat_help report all . > all_report_help

  Additional topics:

    API                         execute
    balance                     experiment
    build                       first_example
    counters                    overview
    demos                       report
    environment                 run

pat_help (.=quit ,=back ^=up /=top ~=search)
=>
```

Using Hardware Performance Counters on the Cray XT

Luiz DeRose
Programming Environment Director
Cray Inc.
ldr@cray.com

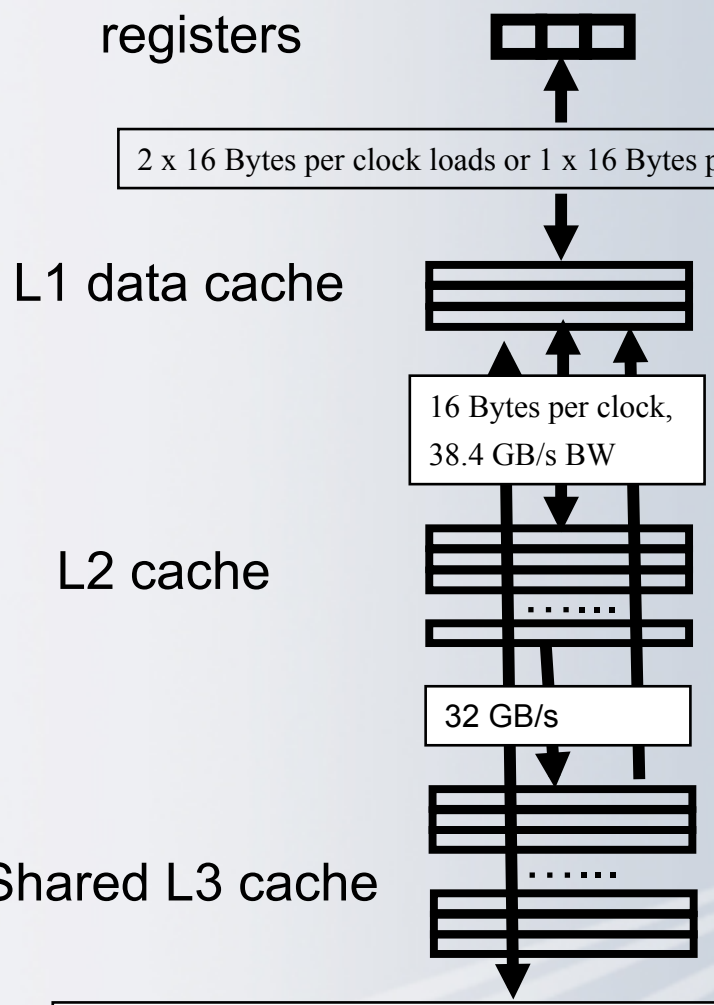
ORNL
October 15, 2008

AMD Opteron Processor



- 36 entry FPU instruction scheduler
- 64-bit/80-bit FP Realized throughput (1 Mul + 1 Add)/cycle: 1.9 FLOPs/cycle
- 32-bit FP Realized throughput (2 Mul + 2 Add)/cycle: 3.4+ FLOPs/cycle

Simplified memory hierachy on the Quad Core AMD Opteron



16 SSE2 128-bit registers
16 64 bit registers

- 64 Byte cache line, 8 banks
- complete data cache lines are loaded from main memory, if not in L2 or L3 cache
- if L1 data cache needs to be refilled, then storing back to L2 cache, if L2 needs to be refilled, storing back to L3
- Items in L1 and L2 are exclusive, L3 is “sharing aware”

- 64 Byte cache line
- write back cache: data offloaded from L1 data cache are stored here first until they are flushed out to main memory

16 Bytes wide => 10.6 GB/s for DDR2-667, 73ns

Main memory

8GB/s over coherent Hyper Transport, 115ns

Remote memory

Hardware Performance Counters

- AMD Opteron Hardware Performance Counters
  - Four 48-bit performance counters.
    - Each counter can monitor a single event
      - Count specific processor events
        - » the processor increments the counter when it detects an occurrence of the event
        - » (e.g., cache misses)
      - Duration of events
        - » the processor counts the number of processor clocks it takes to complete an event
        - » (e.g., the number of clocks it takes to return data from memory after a cache miss)
  - Time Stamp Counters (TSC)
    - Cycles (user time)

PAPI Predefined Events

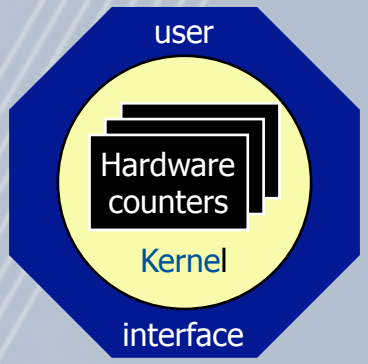
- Common set of events deemed relevant and useful for application performance tuning
  - Accesses to the memory hierarchy, cycle and instruction counts, functional units, pipeline status, etc.
  - The “papi_avail” utility shows which predefined events are available on the system – execute on compute node
- PAPI also provides access to native events
  - The “papi_native_avail” utility lists all AMD native events available on the system – execute on compute node
- Information on PAPI and AMD native events
  - pat_help counters
  - man papi_counters
  - For more information on AMD counters:
    - http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/26049.PDF

Hardware Counters Selection

- PAT_RT_HWPC <set number> | <event list>
  - Specifies hardware counter events to be monitored
    - A set number can be used to select a group of predefined hardware counters events (recommended)
      - CrayPat provides 19 groups on the Cray XT systems
    - Alternatively a list of hardware performance counter event names can be used
      - Maximum of 4 events
    - Both formats can be specified at the same time, with later definitions overriding previous definitions
    - Hardware counter events are not collected by default
    - Hardware counters collection is not supported with sampling on systems running Catamount on the compute nodes

Accuracy Issues

- Granularity of the measured code
  - If not sufficiently large enough, overhead of the counter interfaces may dominate
- Pay attention to what is not measured:
  - Out-of-order processors
  - Speculation
  - Lack of standard on what is counted
    - Microbenchmarks can help determine accuracy of the hardware counters
- For more information on AMD counters:
  - architecture manuals:
    - http://www.amd.com/us-en/assets/content_type/white_papers_and_tech_docs/26049.PDF



Hardware Performance Counters

```
Hardware performance counter events:
   PAPI_TOT_INS        Instructions completed
   PAPI_L1_DCA         Level 1 data cache accesses
   PAPI_FP_OPS         Floating point operations
   DATA_CACHE_MISSES   Data Cache Misses
   CYCLES_USER         User Cycles (approx, from clock ticks)

Estimated minimum overhead per call of a traced function,
   which was subtracted from the data shown in this report
   (for raw data, use the option:  -s overhead=include):
     PAPI_TOT_INS          2021.905  instr
     PAPI_L1_DCA           1275.739  refs
     PAPI_FP_OPS              0.000  ops
     DATA_CACHE_MISSES        7.702  misses
     CYCLES_USER              0.000  cycles
     Time                     2.054  microseconds
```

Hardware Performance Counters

```
==========================================================================
USER / mlwxyz_
--------------------------------------------------------------------------
  Time%                                                  43.3%
  Time                                                5.392446
  Imb.Time                                            0.041565
  Imb.Time%                                              0.8%
  Calls                                                    10
  PAPI_L1_DCM                 19.103M/sec           102483839 misses
  PAPI_TOT_INS               275.051M/sec          1475564421 instr
  PAPI_L1_DCA                170.207M/sec           913107237 refs
  PAPI_FP_OPS                393.744M/sec          2112315940 ops
  User time (approx)           5.365 secs          11265843750 cycles
  Average Time per Call                              0.539245 sec/call
  Cycles                       5.365 secs          11265843750 cycles
  User time (approx)           5.365 secs          11265843750 cycles
  Utilization rate                                       99.5%
  Instr per cycle                                         0.13 inst/cycle
  HW FP Ops / Cycles                                      0.19 ops/cycle
  HW FP Ops / User time      393.744M/sec          2112315940 ops          4.7%peak
  HW FP Ops / WCT            391.718M/sec
  HW FP Ops / Inst                                      143.2%
  Computation intensity                                   2.31 ops/ref
  MIPS                        8801.64M/sec
  MFLOPS                     12599.82M/sec
  Instructions per LD ST                                  1.62 inst/ref
  LD & ST per D1 miss                                     8.91 refs/miss
  D1 cache hit ratio                                     88.8%
  LD ST per Instructions                                 61.9%
```

PAT_RT_HWPC=0
Flat profile data
Hard counts
Derived metrics

PAT_RT_HWPC=1 (Summary with TLB)

```
  PAPI_TLB_DM  Data translation lookaside buffer misses
  PAPI_L1_DCA  Level 1 data cache accesses
  PAPI_FP_OPS  Floating point operations
  DC_MISS      Data Cache Miss
  User_Cycles  Virtual Cycles
==========================================================================
USER / mlwxyz_
--------------------------------------------------------------------------
  Time%                                        43.9%
  Time                                     5.390791
  Imb.Time                                 0.040113
  Imb.Time%                                    0.8%
  Calls                                          10
  PAPI_L1_DCM              19.152M/sec    102649233 misses
  PAPI_TLB_DM               0.302M/sec      1619551 misses
  PAPI_L1_DCA             170.363M/sec    913093302 refs
  PAPI_FP_OPS             394.112M/sec   2112315940 ops
  User time (approx)        5.360 secs  11255343750 cycles
  Average Time per Call                    0.539079 sec/call
  Cycles                    5.360 secs  11255343750 cycles
  User time (approx)        5.360 secs  11255343750 cycles
  Utilization rate                            99.4%
  HW FP Ops / Cycles                           0.19 ops/cycle
  HW FP Ops / User time   394.112M/sec   2112315940 ops         4.7%peak
  HW FP Ops / WCT         391.838M/sec
  Computation intensity                        2.31 ops/ref
  MFLOPS                12611.58M/sec
  LD & ST per TLB miss                       563.79 refs/miss
  LD & ST per D1 miss                          8.90 refs/miss
  D1 cache hit ratio                          88.8%
  % TLB misses / cycle                         0.0%
```

PAT_RT_HWPC=2 (L1 and L2 Metrics)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                  44.3%
  Time                                                5.395990
  Imb.Time                                            0.041032
  Imb.Time%                                               0.8%
  Calls                                                     10
  REQUESTS_TO_L2:DATA                  69.315M/sec     371505133 req
  DATA_CACHE_REFILLS:
    L2_MODIFIED:L2_OWNED:
    L2_EXCLUSIVE:L2_SHARED             14.501M/sec      77720419 fills
  DATA_CACHE_REFILLS_FROM_SYSTEM:
    ALL                                19.011M/sec     101891778 fills
  PAPI_L1_DCA                         170.363M/sec     913093419 refs
  User time (approx)                    5.360 secs   11255343750 cycles
  Average Time per Call                                 0.539599 sec/call
  Cycles                                5.360 secs   11255343750 cycles
  User time (approx)                    5.360 secs   11255343750 cycles
  Utilization rate                                        99.3%
  LD & ST per D1 miss                                      5.08 refs/miss
  D1 cache hit ratio                                      80.3%
  LD & ST per D2 miss                                      8.96 refs/miss
  D2 cache hit ratio                                      43.3%
  D1+D2 cache hit ratio                                   88.8%
  Effective D1+D2 Reuse                                    0.14 refs/byte
  System to D1 refill                  19.011M/sec     101891778 lines
  System to D1 bandwidth             1160.325MB/sec    6521073784 bytes
  L2 to Dcache bandwidth              885.066MB/sec    4974106824 bytes
```

PAT_RT_HWPC=3 (Bandwidth)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                    44.0%
  Time                                                  5.393606
  Imb.Time                                              0.054000
  Imb.Time%                                                1.0%
  Calls                                                      10
  QUADWORDS_WRITTEN_TO_SYSTEM:
    ALL                                 76.516M/sec     410363958 ops
  DATA_CACHE_REFILLS:
    L2_MODIFIED:L2_OWNED:
    L2_EXCLUSIVE:L2_SHARED              14.494M/sec      77731399 fills
  DATA_CACHE_REFILLS_FROM_SYSTEM:
    ALL                                 18.999M/sec     101891701 fills
  DATA_CACHE_LINES_EVICTED:ALL          52.589M/sec     282042348 ops
  User time (approx)                     5.363 secs   11262496875 cycles
  Average Time per Call                                  0.539361 sec/call
  Cycles                                 5.363 secs   11262496875 cycles
  User time (approx)                     5.363 secs   11262496875 cycles
  Utilization rate                                         99.4%
  D2 cache hit ratio                                       43.3%
  System to D1 refill                 18.999M/sec     101891701 lines
  System to D1 bandwidth            1159.587MB/sec    6521068888 bytes
  L2 to Dcache bandwidth             884.629MB/sec    4974809544 bytes
  L2 to System BW per core           583.773MB/sec    3282911662 bytes
================================================================================
```

PAT_RT_HWPC=5 (Floating point mix)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                  44.2%
  Time                                                5.391580
  Imb.Time                                            0.060782
  Imb.Time%                                               1.2%
  Calls                                                     10
  RETIRED_MMX_AND_FP_INSTRUCTIONS:
    PACKED_SSE_AND_SSE2              215.454M/sec   1158535981 instr
  PAPI_FML_INS                       249.283M/sec   1340443200 ops
  PAPI_FAD_INS                       143.546M/sec    771872740 ops
  PAPI_FDV_INS                        33.469M/sec    179967310 ops
  User time (approx)                    5.377 secs  11292093750 cycles
  Average Time per Call                               0.539158 sec/call
  Cycles                                5.377 secs  11292093750 cycles
  User time (approx)                    5.377 secs  11292093750 cycles
  Utilization rate                                        99.7%
  HW FP Ops / Cycles                                       0.19 ops/cycle
  HW FP Ops / User time              392.829M/sec   2112315940 ops        4.7%peak
  HW FP Ops / WCT                    391.780M/sec
  FP Multiply / FP Ops                                    63.5%
  FP Add / FP Ops                                         36.5%
  MFLOPS                           12570.53M/sec
================================================================================
```

PAT_RT_HWPC=6 (Stalls / Resources Idle)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                          44.6%
  Time                                       5.394319
  Imb.Time                                   0.046703
  Imb.Time%                                      0.9%
  Calls                                            10
  INSTRUCTION_FETCH_STALL  2054.169M/sec  10985952007 stalls
  PAPI_FPU_IDL                0.019 secs     40573828 cycles
  PAPI_STL_ICY                0.004 secs      7583186 cycles
  PAPI_RES_STL                5.078 secs  10664161800 cycles
  User time (approx)          5.348 secs  11231062500 cycles
  Average Time per Call                      0.539432 sec/call
  Cycles                      5.348 secs  11231062500 cycles
  User time (approx)          5.348 secs  11231062500 cycles
  Utilization rate                              99.1%
  Total time stalled          5.078 secs  10664161800 cycles    95.0%
  Time I Fetch Stalled        5.231 secs  10985952007 cycles    97.8%
  Avg Time FPUs idle          0.010 secs     20286914 cycles    0.2%
  Time Decoder empty          0.004 secs      7583186 cycles    0.1%
========================================================================
```

PAT_RT_HWPC=7 (Stalls/ Resources Full)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                43.9%
  Time                                              5.390339
  Imb.Time                                          0.044999
  Imb.Time%                                             0.9%
  Calls                                                   10
  DECODER_EMPTY                     0.004 secs       7586977 cycles
  DISPATCH_STALLS                1987.284M/sec  10655566763 stalls
  DISPATCH_STALL_FOR_FPU_FULL     690.689M/sec   3703389803 stalls
  DISPATCH_STALL_FOR_LS_FULL     1036.705M/sec   5558685175 stalls
  User time (approx)                5.362 secs  11259937500 cycles
  Average Time per Call                           0.539034 sec/call
  Cycles                            5.362 secs  11259937500 cycles
  User time (approx)                5.362 secs  11259937500 cycles
  Utilization rate                                     99.5%
  Total time stalled                5.074 secs  10655566763 cycles    94.6%
  Avg Time FPUs stalled             0.882 secs   1851694902 cycles    16.4%
  Avg Time LSs stalled              1.323 secs   2779342588 cycles    24.7%
  Time Decoder empty                0.004 secs      7586977 cycles    0.1%
================================================================================
```

PAT_RT_HWPC 8 & 9

```
-------------------------------------------------------------------
Set 8: Branches
  PAPI_BR_TKN   Conditional branch instructions taken
  PAPI_BR_MSP   Conditional branch instructions mispredicted
  PAPI_TOT_INS  Instructions completed
  IC_MISS       IC Miss
  User_Cycles   Virtual Cycles


-------------------------------------------------------------------

Set 9: Instructions
  PAPI_L2_ICM   Level 2 instruction cache misses
  PAPI_L1_ICA   Level 1 instruction cache accesses
  IC_MISS       IC Miss
  IC_L2_REFILL  Refill from L2
  User_Cycles   Virtual Cycles
```

PAT_RT_HWPC=10 (Cache Hierarchy)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                          43.8%
  Time                                       5.398147
  Imb.Time                                   0.047806
  Imb.Time%                                      0.9%
  Calls                                            10
  PAPI_L1_DCM              19.132M/sec    102761682 misses
  PAPI_L2_DCM               6.377M/sec     34250616 misses
  PAPI_L3_TCM              57.550M/sec    309115904 misses
  PAPI_L1_DCA             169.998M/sec    913104078 refs
  User time (approx)        5.371 secs  11279625000 cycles
  Average Time per Call                    0.539815 sec/call
  Cycles                    5.371 secs  11279625000 cycles
  User time (approx)        5.371 secs  11279625000 cycles
  Utilization rate                            99.5%
  LD & ST per D1 miss                          8.89 refs/miss
  D1 cache hit ratio                          88.7%
  LD & ST per D2 miss                         26.66 refs/miss
  D2 cache hit ratio                          66.7%
  D1+D2 cache hit ratio                       96.2%
  Effective D1+D2 Reuse                        0.42 refs/byte
================================================================================
```

PAT_RT_HWPC=11 (Quad Core FLOPS)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                  44.1%
  Time                                                5.387697
  Imb.Time                                            0.048925
  Imb.Time%                                               0.9%
  Calls                                                     10
  RETIRED_SSE_OPERATIONS:
    SINGLE_ADD_SUB_OPS:
    SINGLE_MUL_OPS:OP_TYPE          395.125M/sec   2112315940 ops
  RETIRED_SSE_OPERATIONS:
    SINGLE_DIV_OPS:OP_TYPE           33.664M/sec    179967310 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_ADD_SUB_OPS:
    DOUBLE_MUL_OPS:OP_TYPE                                 0 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_DIV_OPS:OP_TYPE                                 0 ops
  User time (approx)                 5.346 secs  11226468750 cycles
  Average Time per Call                              0.538770 sec/call
  Cycles                             5.346 secs  11226468750 cycles
  User time (approx)                 5.346 secs  11226468750 cycles
  Utilization rate                                      99.2%
  HW FP Ops / Cycles                                     0.20 ops/cycle
  HW FP Ops / User time            428.790M/sec   2292283250 ops        5.1%peak
  HW FP Ops / WCT                  425.466M/sec
  Weighted HW FP Ops / Cycles                            0.27 ops/cycle
  Weighted HW FP Ops / User time   563.447M/sec   3012152490 ops        5.1%peak
  Weighted HW FP Ops / WCT         559.080M/sec
  MFLOPS                         13721.27M/sec
================================================================================
```

PAT_RT_HWPC=12 (QC Vectorization)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                          43.5%
  Time                                       5.387259
  Imb.Time                                   0.059600
  Imb.Time%                                      1.1%
  Calls                                            10
  RETIRED_SSE_OPERATIONS:
    SINGLE_ADD_SUB_OPS:
    SINGLE_MUL_OPS          104.066M/sec    558216100 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_ADD_SUB_OPS:
    DOUBLE_MUL_OPS                                 0 ops
  RETIRED_SSE_OPERATIONS:
    SINGLE_ADD_SUB_OPS:
    SINGLE_MUL_OPS:OP_TYPE  393.790M/sec   2112315940 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_ADD_SUB_OPS:
    DOUBLE_MUL_OPS:OP_TYPE                         0 ops
  User time (approx)          5.364 secs  11264531250 cycles
  Average Time per Call                      0.538726 sec/call
  Cycles                      5.364 secs  11264531250 cycles
  User time (approx)          5.364 secs  11264531250 cycles
  Utilization rate                               99.6%
========================================================================
```

Vectorization Example

```
  PAPI_FML_INS              Floating point multiply instructions
  PAPI_FAD_INS              Floating point add instructions
  FR_FPU_SSE_SSE2_PACKED    Retired FPU instructions - Combined packed SSE and SSE2 instructions
  FR_FPU_SSE_SSE2_SCALAR    Retired FPU instructions - Combined scalar SSE and SSE2 instructions
  User_Cycles               Virtual Cycles
=============================================================================
USER / sweep_
-----------------------------------------------------------------------------
  Time%                                          97.5%
  Time                                       3.230243
  Imb.Time                                   0.102413
  Imb.Time%                                      3.1%
  Calls                                           576
  PAPI_FML_INS            14751.121M/sec   47682418308 instr
  PAPI_FAD_INS            16874.595M/sec   54546466150 instr
  FR_FPU_SSE_SSE2_PACKED                             0 instr
  FR_FPU_SSE_SSE2_SCALAR  43000.358M/sec  138996966424 instr
  User time                   3.232 secs    7757905716 cycles
  Utilization rate                             100.0%
  HW FP Ops / Cycles                             13.18 ops/cycle
  HW FP Ops / User time   31625.716M/sec  102228884458 ops        13.7%peak
  HW FP Ops / WCT         31625.716M/sec
  FP Multiply / FP Ops                           46.6%
  FP Add / FP Ops                                53.4%

When compiled with fastsse:
=============================================================================
USER / sweep_
-----------------------------------------------------------------------------
  Time%                                          97.0%
  Time                                       2.577571
  Imb.Time                                   0.101843
  Imb.Time%                                      3.9%
  Calls                                           576
  PAPI_FML_INS            16061.952M/sec   41438628312 instr
  PAPI_FAD_INS            18681.139M/sec   48195934483 instr
  FR_FPU_SSE_SSE2_PACKED   9315.154M/sec   24032397312 instr
  FR_FPU_SSE_SSE2_SCALAR  39220.314M/sec  101185461233 instr
  User time                   2.580 secs    6191819596 cycles
  Utilization rate                             100.0%
  HW FP Ops / Cycles                             14.48 ops/cycle
  HW FP Ops / User time   34743.091M/sec   89634562795 ops        15.1%peak
  HW FP Ops / WCT         34743.091M/sec
  FP Multiply / FP Ops                           46.2%
  FP Add / FP Ops                                53.8%
```

PAT_RT_HWPC=13 (FLOPS Single Prec.)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                              44.3%
  Time                                             5.404342
  Imb.Time                                         0.056660
  Imb.Time%                                            1.1%
  Calls                                                  10
  RETIRED_SSE_OPERATIONS:
    SINGLE_ADD_SUB_OPS           38.067M/sec     204881380 ops
  RETIRED_SSE_OPERATIONS:
    SINGLE_MUL_OPS               65.649M/sec     353334720 ops
  RETIRED_SSE_OPERATIONS:
    SINGLE_ADD_SUB_OPS:OP_TYPE  143.412M/sec     771872740 ops
  RETIRED_SSE_OPERATIONS:
    SINGLE_MUL_OPS:OP_TYPE      249.052M/sec    1340443200 ops
  User time (approx)              5.382 secs   11302593750 cycles
  Average Time per Call                          0.540434 sec/call
  Cycles                          5.382 secs   11302593750 cycles
  User time (approx)              5.382 secs   11302593750 cycles
  Utilization rate                                   99.6%
========================================================================
```

PAT_RT_HWPC=14 (FLOPS Double Prec.)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                              44.4%
  Time                                             5.393922
  Imb.Time                                         0.045698
  Imb.Time%                                            0.9%
  Calls                                                  10
  RETIRED_SSE_OPERATIONS:
    DOUBLE_ADD_SUB_OPS                                    0 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_MUL_OPS                                        0 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_ADD_SUB_OPS:OP_TYPE                            0 ops
  RETIRED_SSE_OPERATIONS:
    DOUBLE_MUL_OPS:OP_TYPE                                0 ops
  User time (approx)            5.342 secs  11217215625 cycles
  Average Time per Call                          0.539392 sec/call
  Cycles                        5.342 secs  11217215625 cycles
  User time (approx)            5.342 secs  11217215625 cycles
  Utilization rate                                  99.0%
========================================================================
```

PAT_RT_HWPC=15 (L3 socket level)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                   44.1%
  Time                                                 5.397331
  Imb.Time                                             0.058790
  Imb.Time%                                               1.1%
  Calls                                                     10
  L3_EVICTIONS:ALL                    96.567M/sec    517503487 ops
  READ_REQUEST_TO_L3_CACHE:ALL       114.153M/sec    611745028 req
  L3_CACHE_MISSES:ALL                 57.352M/sec    307348659 misses
  L3_FILLS_CAUSED_BY_L2_EVICTIONS:
    ALL                               21.843M/sec    117055235 fills
  User time (approx)                   5.359 secs  11253900000 cycles
  Average Time per Call                                0.539733 sec/call
  Cycles                               5.359 secs  11253900000 cycles
  User time (approx)                   5.359 secs  11253900000 cycles
  Utilization rate                                        99.3%
  L3 cache hit ratio                                      49.8%
================================================================================
```

PAT_RT_HWPC=16 (L3 core level reads)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                          43.5%
  Time                                       5.396417
  Imb.Time                                   0.038339
  Imb.Time%                                      0.7%
  Calls                                            10
  READ_REQUEST_TO_L3_CACHE:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_0_SELECT            114.016M/sec    611411065 req
  READ_REQUEST_TO_L3_CACHE:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_1_SELECT            114.016M/sec    611411065 req
  READ_REQUEST_TO_L3_CACHE:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_2_SELECT            114.016M/sec    611411065 req
  READ_REQUEST_TO_L3_CACHE:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_3_SELECT            114.016M/sec    611411065 req
  User time (approx)           5.362 secs  11261250000 cycles
  Average Time per Call                       0.539642 sec/call
  Cycles                       5.362 secs  11261250000 cycles
  User time (approx)           5.362 secs  11261250000 cycles
  Utilization rate                               99.4%
========================================================================
```

PAT_RT_HWPC=17 (L3 core level misses)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                          44.4%
  Time                                       5.405786
  Imb.Time                                   0.033002
  Imb.Time%                                      0.6%
  Calls                                            10
  L3_CACHE_MISSES:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_0_SELECT          38.708M/sec    207924389 misses
  L3_CACHE_MISSES:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_1_SELECT          38.708M/sec    207924389 misses
  L3_CACHE_MISSES:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_2_SELECT          44.973M/sec    241574844 misses
  L3_CACHE_MISSES:
    READ_BLOCK_MODIFY:
    READ_BLOCK_EXCLUSIVE:
    READ_BLOCK_SHARED:
    CORE_3_SELECT          44.973M/sec    241574844 misses
  User time (approx)        5.372 secs  11280281250 cycles
  Average Time per Call                    0.540579 sec/call
  Cycles                    5.372 secs  11280281250 cycles
  User time (approx)        5.372 secs  11280281250 cycles
  Utilization rate                             99.4%
========================================================================
```

PAT_RT_HWPC=18 (L3 core fills from L2 Evictions)

```
========================================================================
USER / mlwxyz_
------------------------------------------------------------------------
  Time%                                                  44.1%
  Time                                                5.390285
  Imb.Time                                            0.044635
  Imb.Time%                                               0.8%
  Calls                                                     10
  L3_FILLS_CAUSED_BY_L2_EVICTIONS:
    MODIFIED:OWNED:EXCLUSIVE:
    SHARED:CORE_0_SELECT          21.842M/sec    117136467 fills
  L3_FILLS_CAUSED_BY_L2_EVICTIONS:
    MODIFIED:OWNED:EXCLUSIVE:
    SHARED:CORE_1_SELECT          21.842M/sec    117136467 fills
  L3_FILLS_CAUSED_BY_L2_EVICTIONS:
    MODIFIED:OWNED:EXCLUSIVE:
    SHARED:CORE_2_SELECT          21.842M/sec    117136467 fills
  L3_FILLS_CAUSED_BY_L2_EVICTIONS:
    MODIFIED:OWNED:EXCLUSIVE:
    SHARED:CORE_3_SELECT          21.842M/sec    117136467 fills
  User time (approx)               5.363 secs  11261906250 cycles
  Average Time per Call                          0.539029 sec/call
  Cycles                           5.363 secs  11261906250 cycles
  User time (approx)               5.363 secs  11261906250 cycles
  Utilization rate                                    99.5%
========================================================================
```

PAT_RT_HWPC=19 (Prefetchs)

```
================================================================================
USER / mlwxyz_
--------------------------------------------------------------------------------
  Time%                                                  44.1%
  Time                                                5.402184
  Imb.Time                                            0.059770
  Imb.Time%                                               1.1%
  Calls                                                     10
  PREFETCH_INSTRUCTIONS_DISPATCHED:
    LOAD                                 6.484M/sec     34905828 refs
  DATA_PREFETCHES:CANCELLED                                 0 refs
  PREFETCH_INSTRUCTIONS_DISPATCHED:
    STORE                                0.741M/sec      3988078 refs
  DATA_PREFETCHES:ATTEMPTED                                 0 refs
  User time (approx)                     5.383 secs  11305153125 cycles
  Average Time per Call                                0.540218 sec/call
  Cycles                                 5.383 secs  11305153125 cycles
  User time (approx)                     5.383 secs  11305153125 cycles
  Utilization rate                                        99.7%
================================================================================
```

Performance Measurement and Visualization on the Cray XT

Questions / Comments

Thank You!

ORNL

October 15, 2008